K. J. Somaiya Institute of Technology, Sion, Mumbai
(An Autonomous Institute Affiliated to the University of Mumbai)

End Semester Examination: May - June 2025
Program: B.Tech. (Information Technology) Scheme: IIB
Regular Examination: TY Semester: VI
Course Code: Data Mining and Business Intelligence Course Name: ITC601
Date of Exam: 20/05/2025 Duration: 2.5 Hours Max. Marks: 60

Instructions:
(1) All questions are compulsory.
(2) Draw neat diagrams wherever applicable.
(3) Assume suitable data, if necessary.

<table>
<tr><th>Ques. No.</th><th>Question</th><th>Max. Marks</th><th>CO</th><th>BT Level</th></tr>
<tr><td>Q1</td><td>Solve <u>any two</u> questions out of three: (05 marks each)</td><td rowspan="4">10</td><td></td><td></td></tr>
<tr><td>a)</td><td>Consider the Confusion Matrix given below for two classes: Churn and Retain in an application identifying customers likely to cancel subscriptions. Calculate the Accuracy, Error, Precision, and Recall.
<table>
<tr><th></th><th>Churn</th><th>Retain</th></tr>
<tr><td>Churn</td><td>600</td><td>150</td></tr>
<tr><td>Retain</td><td>250</td><td>1000</td></tr>
</table></td><td>3</td><td>A</td></tr>
<tr><td>b)</td><td>Differentiate classification and clustering by analyzing their use in solving a business or scientific problem.</td><td>4</td><td>A</td></tr>
<tr><td>c)</td><td>Explain the kinds of patterns that can be mined.</td><td>1</td><td>U</td></tr>
<tr><td>Q2</td><td>Solve <u>any two</u> questions out of three: (05 marks each)</td><td rowspan="4">10</td><td></td><td></td></tr>
<tr><td>a)</td><td>Sketch a scatter plot for data of (Age, Weight) recorded by a Paediatrician: (7, 9), (8, 10), (9, 12), (10, 14), (11, 18), (12, 22), (12, 24), (13, 23).</td><td>2</td><td>A</td></tr>
<tr><td>b)</td><td>Explain Decision Support Systems in Business Intelligence.</td><td>6</td><td>U</td></tr>
<tr><td>c)</td><td>Given a dataset of 5 points, briefly describe how agglomerative and divisive clustering would approach the formation of clusters.</td><td>5</td><td>A</td></tr>
<tr><td>Q3</td><td>Solve <u>any two</u> questions out of three. (10 marks each)</td><td rowspan="2">20</td><td></td><td></td></tr>
<tr><td>a)</td><td>Consider a medical symptom dataset in the form <PatientID: Symptoms> as:
<P1: Fever, Cough, Fatigue>,
<P2: Cough, Headache, Fatigue>,
<P3: Fever, Cough>,
<P4: Cough, Fatigue>,
<P5: Fever, Cough, Headache>.
If the minimum support is 20% and the minimum confidence is 70%, apply the Apriori algorithm to find frequent symptom combinations and generate strong association rules.</td><td>5</td><td>A</td></tr>
</table>

K. J. Somaiya Institute of Technology, Sion, Mumbai
(An Autonomous Institute Affiliated to the University of Mumbai)

End Semester Examination: May - June 2025
Program: B.Tech. (Information Technology) Scheme: IIB
Regular Examination: TY Semester: VI
Course Code: Data Mining and Business Intelligence Course Name: ITC601
Date of Exam: 20/05/2025 Duration: 2.5 Hours Max. Marks: 60

<table>
<tr><td>b)</td><td>Explain how OLAP operations such as roll-up, drill-down, slice, and dice can be applied to analyze retail sales data for decision-making.</td><td></td><td>1</td><td>U</td></tr>
<tr><td>c)</td><td>Suppose that the data mining task is to cluster the coordinates of disaster relief camps for optimal resource allocation. The locations are: D1(3, 12), D2(4, 7), D3(10, 6), E1(6, 10), E2(9, 7), E3(7, 5), F1(2, 3), F2(5, 11). Each point represents a camp's location (x, y). Initially, take D1, E1, and F1 as the cluster centers. Use Euclidean distance and apply one iteration of the k-means algorithm to assign clusters and compute the new cluster centers.</td><td></td><td>4</td><td>A</td></tr>
<tr><td>Q4</td><td>Solve <u>any two</u> questions out of three. (10 marks each)</td><td rowspan="4">20</td><td></td><td></td></tr>
<tr><td>a)</td><td>Consider data of monthly advertising expenses and monthly sales revenue for a retail business:
<table>
<tr><th>Month</th><th>Advertising Expense (in INR)</th><th>Sales Revenue (in INR)</th></tr>
<tr><td>January</td><td>5000</td><td>60000</td></tr>
<tr><td>February</td><td>6000</td><td>65000</td></tr>
<tr><td>March</td><td>7000</td><td>70000</td></tr>
<tr><td>April</td><td>5500</td><td>62000</td></tr>
<tr><td>May</td><td>7500</td><td>72000</td></tr>
</table>
Perform Correlation analysis.</td><td>2</td><td>A</td></tr>
<tr><td>b)</td><td>Apply the Decision Tree algorithm to compute Information Gain and select the best root node.
<table>
<tr><th>Income</th><th>Gender</th><th>Education</th><th>Purchase Interest</th></tr>
<tr><td>Low</td><td>Male</td><td>Elementary</td><td>Negative</td></tr>
<tr><td>Medium</td><td>Male</td><td>Intermediate</td><td>Negative</td></tr>
<tr><td>High</td><td>Female</td><td>Intermediate</td><td>Positive</td></tr>
<tr><td>Low</td><td>Male</td><td>Complete</td><td>Positive</td></tr>
<tr><td>Medium</td><td>Female</td><td>Elementary</td><td>Positive</td></tr>
<tr><td>High</td><td>Male</td><td>Complete</td><td>Positive</td></tr>
<tr><td>Low</td><td>Female</td><td>Intermediate</td><td>Negative</td></tr>
<tr><td>Low</td><td>Male</td><td>Intermediate</td><td>Negative</td></tr>
<tr><td>Medium</td><td>Female</td><td>Complete</td><td>Positive</td></tr>
</table></td><td>3</td><td>A</td></tr>
<tr><td>c)</td><td>Consider the case of credit card fraud detection. Apply each phase of the KDD process on transaction data to derive Business Intelligence.</td><td>6</td><td>A</td></tr>
</table>

****